

**ECON742: Exercise 5**  
**Regression Discontinuity Design**

1. The data comes from the paper “Politics, Markets and Schools: Quasi-Experimental Evidence on the Impact of Autonomy and Competition from a Truly Revolutionary UK Reform” from Damon Clark, downloadable at the following address

<http://www.nber.org/~confer/2005/si2005/ch/clark.pdf>.

The database includes the following variables:

- `passrate0` : the pass rate of pupils in the school in the year immediately prior to the vote
- `passrate2` : the pass rate of pupils in the school two years after the vote
- `dpass`: the change in the pass rate = `passrate2`-`passrate0`
- `vote`: the percentage vote in favour of the GM status
- `win`: a dummy variable if the vote was more than 50%
- `win_vote` =  $\text{win} \cdot (50 - \text{vote}) / 100$
- `win_vote_2` =  $\text{win} \cdot (\text{vote squared})$
- `lose_vote` =  $\text{lose} \cdot (50 - \text{vote}) / 100$
- `lose_vote_2` =  $\text{lose} \cdot (\text{vote squared})$

1. (a) Graph `dpass` as a function of `vote`. What do you see? Is the discontinuity clear?
- (b) Reproduce the results of Table 3a (without using the weights, not available in the database). What column do you prefer? Why does the author do a regression of `dpass` on `win` and `vote`, then a regression of `dpass` on `win`, `lose_vote` and `win_vote`? Could we have taken as the definition of `lose_vote` the multiplication of `lose` by `vote`?
- (c) Why do the authors only keep the schools where the votes were between 15% and 85%? What is the change if you keep all the schools?
- (d) Instead of using `dpass` as a dependent variable, use `passrate2`. What are the results? What does it mean in terms of regression discontinuity design?
- (e) Regres `passrate0` on `win` and `vote`. How to interpret the coefficient of `win`? What does it mean? Is the estimation correct?